**Stat and Data Analysis – FINAL CLASS PROJECT**

**Purpose:**

To demonstrate knowledge and skills learned throughout the semester in Statistics and Data Analysis.

**Objective:**

Given a data set, students will need to create a random sample, find summary statistics, create graphs with analysis, and run inference tests about the population from the sample.

***Students will work in groups of 2 for this project. Each member of the group is expected to participate in the completion and presentation of the project.***

**Background:**

A college professor of an introductory statistics class has been keeping records of his past students for the last few semesters. He has recorded many variables about his students including gender, major, midterm exam grade, and final exam grade. He is interested in doing some data analysis of his students and has asked your group to complete the work for him. He has given you data on 600 of his students. He has six major questions that he wants answered:

1. Are the final exam scores averaging less than a 75%?
2. Do more males take this class than females?
3. Are the final exam scores the same for males and females?
4. Is there a relationship between midterm scores and final exam scores?
5. Do students tend to do worse on the final exam than the midterm exam?
6. Are the choices of majors uniformly distributed?

**Technology**

You will be using the laptops and the programs Fathom, Excel, and Powerpoint. You will use Fathom to construct all graphs and summary statistics needed. Bar graphs will be done on excel. You will create a power point with your results.

**Dates**

* You will have 4 days in class to work.
* The project will be due BEFORE CLASS on the due date.
  + Power point, Excel, & Fathom document dropped in my folder or emailed to me ***before class***
  + Printed power point slides (6 to a page) must be handed in at the ***beginning*** of class.

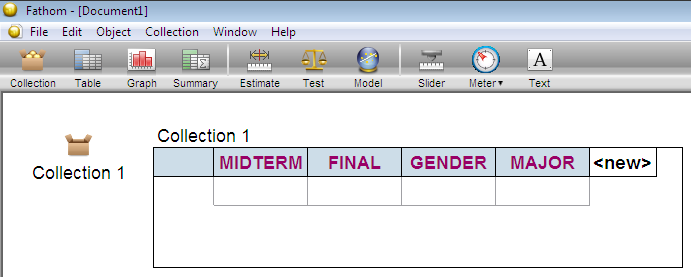
NOTES:

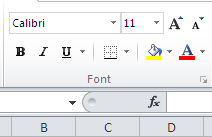
* For all Confidence Intervals & Tests of Significance, you must check conditions and show full work (with full sentences/conclusions).
* For all graphs, describe them in complete sentences and be sure to have units in your descriptions. You MAY copy graphs from Fathom onto your power point.
* For all summary statistics, you may NOT copy summary tables from Fathom. You must type the statistics into your power point.

**Instructions:**

First, create a random sample from the data set (Excel document) you have been given (in my share folder, called Final Project Data) by doing the following:

* Partner #1 should have the data set (Excel document) open on their computer.
* Partner #2 should have Fathom open on their computer, and should have a table open like the one below





* Partner #1: Using randInt(1, 600) on the calculator, randomly select 58 people for your sample. As you select each person, highlight the row in the excel document of the person you selected, & then highlight the row in YELLOW by clicking on the fill button at the top of the screen.

Then read the data for that person to your partner. They will input the data in Fathom.

* Partner #2: As your partner selects and highlights each person, they will read their data to you (midterm, final, gender, major). Input this into your Fathom table.
* Save the excel document on your student drive (or flash drive) with your name on it (example: final project data- smith & johnson). Close the excel document.

***Now you can start your power point and your data analysis.***

* On your power point, put a title slide, and an introduction slide (see the first page of these instructions for the basic idea of the project)
* On the next slide, write clear instructions on how you found your sample.
* **Using your sample**, answer all of the following questions on the power point:

1. Are the final exam scores averaging less than a 75%?

* Create a histogram of the final exam scores
* Create a boxplot of the scores
* Find summary statistics of the scores
* Complete a 1.5\*IQR outlier test
* Describe the distribution (shape, center, spread) based on the histogram and appropriate summary statistics
* Test to see if the ***average*** final exam scores are less than the claim of 75%.

Use α = 0.05. (Be sure to write down your “important info,” like , s, n, df)

* Create and interpret a 90% confidence interval to estimate the true ***average*** final exam score.

1. Do more males take this class than females?

* Create a table of gender. Type this table into Excel. Put the excel table into your power point.
* Create a bar charts of gender (use Excel)
* Find the sample proportions () of males ***and*** of females
* Using the () of males, test if the true ***percent*** of males is 50% or if it is greater than 50%. (Be sure to write down your important info again)
* Construct and interpret a 90% confidence interval to estimate the true ***percent*** of males.

1. Are the final exam scores the same for males and females?

* Create parallel boxplots for final exam scores separated by Gender
* Find summary statistics for the final exam scores separated by Gender
* Describe and COMPARE the two distributions using the graph and appropriate summary statistics (shape, center, spread for each gender & compare each!)
* Make a conclusion: Are the final exam scores the same for males and females??

1. Is there a relationship between midterm scores and final exam scores?

* Create a scatterplot of Midterm scores vs. Final Exam scores
* Describe the scatterplot (form, direction, strength)
* Add the Least Squares Regression Line (LSRL) to the plot (copy this separately onto your powerpoint)
* Find the LSRL using Fathom (use a MODEL) & write this on your powerpoint (don’t forget the hat!)
* Interpret the slope of the LSRL
* Find the correlation(r) and r2
* Interpret r2
* Create a residual plot and interpret it
* Randomly select 2 students from your sample and find their predicted final exam scores. Show your calculations.
* Using the same two students, find the residuals for these predictions
* How well do you think a Midterm grades predict Final Exam grades? Answer this using your correlation, residual plot, and the 2 residuals you calculated.

1. Do students tend to do worse on the final exam than the midterm exam?

* Find the difference of each student’s Final Exam and Midterm Exam (Final – Midterm)
  + *Note: See the “Fathom & Excel Quick Reference” for instructions to help you*
* Create a histogram of these differences
* Find the summary statistics
* Describe the distribution of differences using the graph and appropriate summary statistics (shape, center, spread)
* Conduct an inference test (t-test) on the differences using the following hypotheses:

1. Are majors evenly distributed?

* Create a table of “Choice of major” in Fathom. Transfer this to Excel. Copy the Excel table to your power point.
* Create a bar chart of majors (in Excel).
* Calculate the # and % of students in each major.
* Conduct an inference test to answer the question. (Chi Square Goodness of Fit test) Show all important information! (observed & expected #’s for all categories)

1. Is there an association between sex and choice of major?

* Create a two-way table of sex and choice of major in Fathom. Transfer this to Excel. Copy the Excel table to your power point.
* Create a segmented bar chart on Excel of majors separated by sex. Put major on the x-axis, broken down by gender in each bar.
  + Calculate the % of males and females in each major.
  + Describe the bar chart
* Conduct an inference test to answer the question. You may need to increase your sample size to meet the conditions. (Chi-Square test of independence)

Show all important information! (observed & expected tables)

* You should also include an introduction slide and a conclusion slide. The conclusion slide should summarize your results from ***EACH*** of the numbered analyses above.
* Your power point, excel, and fathom files should be dropped into my drop folder BEFORE class on the due date.
* You should print out your slides (6 to a page) and hand them in at the beginning of class on the due date.

RUBRIC

**ITEM POINTS DEDUCTED**

1. Random Sample
   1. Description of sampling technique \_\_\_\_ 5
2. Are the final exam scores averaging less than a 75%?
   1. Histogram & boxplot \_\_\_\_ 4
   2. Summary Statistics \_\_\_\_ 3
   3. Outlier test \_\_\_\_ 3
   4. Appropriate description (Shape, Center, Spread, in sentences, w/ units) \_\_\_\_ 5
   5. Inference Test \_\_\_\_ 16
      1. Hypotheses(2)
      2. Conditions(3)
      3. Test Statistic(3)
      4. P-Value(4)
      5. Conclusion(4)
   6. Confidence interval \_\_\_\_ 5
      1. Equation with interval(3)
      2. Conclusion(2)
3. Do more males take this class than females?
   1. Excel table \_\_\_\_ 1
   2. Bar Chart (from Excel) \_\_\_\_ 2
   3. Sample Proportions of males and females \_\_\_\_ 2
   4. Inference Test \_\_\_\_ 15
      1. Hypotheses(2)
      2. Conditions(3)
      3. Test Statistic(3)
      4. P-Value(3)
      5. Conclusion(4)
   5. Confidence interval \_\_\_\_ 5
      1. Equation with interval(3)
      2. Conclusion(2)
4. Are the final exam scores the same for males and females?
   1. Parallel Boxplots \_\_\_\_ 2
   2. Summary Statistics of both genders \_\_\_\_ 6
   3. Appropriate comparisons(Shape, Center, Spread) between the genders \_\_\_\_ 6
   4. Conclusion \_\_\_\_ 2
5. Is there a relationship between midterm scores and final exam scores?
   1. Scatterplot, and separate one with LSRL on it \_\_\_\_ 3
   2. Description (form, direction, strength) \_\_\_\_ 3
   3. LSRL given (use MODEL to find eqn, not the line on the scatterplot) \_\_\_\_ 3
   4. Slope interpretation \_\_\_\_ 2
   5. Correlation and r2 given \_\_\_\_ 2
   6. r2 interpreted \_\_\_\_ 2
   7. Residual plot and interpretation \_\_\_\_ 4
   8. 2 students predictions (work shown) \_\_\_\_ 6
   9. Residuals for 2 students predictions (work shown) \_\_\_\_ 2
   10. Strength and appropriateness of linear model (with justification) \_\_\_\_ 3
6. Do students tend to do worse on the final exam than the midterm exam?
   1. Histogram of differences \_\_\_\_ 3
   2. Summary Statistics of differences \_\_\_\_ 3
   3. Appropriate description (Shape, Center, Spread) of histogram \_\_\_\_ 5
   4. Inference Test on differences \_\_\_\_ 16
      1. Hypotheses(1)
      2. Conditions(3)
      3. Test Statistic(4)
      4. P-Value(4)
      5. Conclusion(4)
7. Are majors evenly distributed?
   1. Table of majors (from Excel) \_\_\_\_ 1
   2. Bar chart with # and % of each major \_\_\_\_ 6
   3. Inference Test \_\_\_\_ 20
      1. Observed/Expected tables (2)
      2. Hypotheses(3)
      3. Conditions(2)
      4. Test Statistic(5)
      5. P-Value(4)
      6. Conclusion(4)
8. Is there an association between gender & major?
   1. Two-way table of gender vs. majors (from Excel) \_\_\_\_ 1
   2. Segmented bar chart (from Excel) \_\_\_\_ 2
   3. % of each gender in each major \_\_\_\_ 4
   4. Describe bar chart \_\_\_\_ 3
   5. Inference Test \_\_\_\_ 20
      1. Observed/Expected charts (2)
      2. Hypotheses(3)
      3. Conditions(2)
      4. Test Statistic(5)
      5. P-Value(4)
      6. Conclusion(4)
9. Introduction (2) & Conclusion (7) slides \_\_\_\_ 9
10. Powerpoint, Excel, and Fathom File Dropped into Folder \_\_\_\_ 5
11. Project emailed/dropped on time (-20 per day) \_\_\_\_
12. Individual contribution \_\_\_\_ 20
13. Slides printed, 6 per page \_\_\_\_ 5

TOTAL: \_\_\_\_\_\_ / 235